Prepared for submission to JHEP

# Finding viable Models in SUSY Parameter Spaces with Signal Specific Discovery Potential

**Thomas Burgess**[a] **Jan Øye Lindroos**[a] **Anna Lipniacka**[a] **Heidi Sandaker**[a]

[a] *Department of Physics and Technology, University of Bergen, Norway*

*E-mail:* thomas.burgess@ift.uib.no, jan.lindroos@ift.uib.no, heidi.sandaker@ift.uib.no, anna.lipniacka@ift.uib.no

arXiv:1210.7020v1 [hep-ph] 25 Oct 2012

Abstract: Recent results from ATLAS and CMS indicating a Higgs mass around 125 GeV, further constrain already highly constrained supersymmetric models such as pMSSM or CMSSM/mSUGRA. Finding potentially discoverable and non-excluded regions of model parameter space is becoming increasingly difficult. Several groups have invested large effort in studing the consequences of Higgs mass bounds, upper limits on rare $B$-meson decays, and limits on relic dark matter density on constrained models, aiming on predicting superpartner masses, and establishing likelihood of SUSY models compared to that of the Standard Model vis-avis experimental data. In this paper a general framework for efficient search for discoverable, non-excluded regions of different SUSY spaces giving specific experimental signature of interest is presented. The method employs an improved Markov Chain Monte Carlo scheme exploiting an iteratively updated likelihood function to guide search for viable models. Existing experimental and theoretical bounds as well as the LHC discovery potential are taken into account. This includes recent bounds on relic dark matter density, the Higgs sector and rare $B$-mesons decays. A clustering algorithm is applied to classify selected models according to expected phenomenology enabling automated choice of experimental benchmarks and regions to be used for optimizing searches. The aim is to provide experimentalist with a viable tool helping to target experimental signatures to search for, once a class of models of interest is established. As an example a search for viable CMSSM models with $\tau$-lepton signatures observable with the 2012 LHC data set is presented. In the search 284496 unique models were probed. From these, nine reference benchmark points covering different ranges of phenomenological observables at the LHC were selected.

Keywords: SUSY, CMSSM , MCMC, LHC

## Contents

## 1 Introduction

Supersymmetry (SUSY) may alleviate many of the problems associated with the Standard Model of particle physics (SM) if the mass of the superpartners lays close to the TeV-scale [1], [2]. Furthermore, it provides a natural dark matter candidate in the form of the Lightest Supersymmetric Particle (LSP), if $R$-parity is conserved [3]. However, even the simplest SUSY extension of the SM, the so called Minimal Supersymmetric Standard Model (MSSM), introduces over 100 new free parameters making them very difficult to experimentally constrain. On the other hand a large part of the MSSM parameter space is already ruled out, as it would lead to unobserved phenomena like non-conservation of lepton numbers, flavour changing neutral currents or large CP violation [4]. It is therefore common practice to look at constrained models that assume a partial unification of parameters at some high energy scale and where the dynamics of the high energy theory ensures more viable phenomenologies [5]. The minimal SUper GRAvity model (mSUGRA) [6] is an example of such a constrained model where the SUSY parameters are assumed to unify at the GUT scale into five universal parameters, a common scalar mass $m_0$, a common gaugino mass $m_{1/2}$, the ratio between the SUSY Higgs vacuum expectation values $\tan\beta$, a common trilinear Higgs-sfermion coupling $A_0$, and the sign of the Higgsino mass parameter $\mu$. In the "lighter" version of it, the so called constrained MSSM (CMSSM) [7–9] gravitino mass is not forced to unify at the same scale as other gaugino masses. In NUHM (Non-Universal Higgs Masses) [10, 11] models, masses of Higgs bosons do not unify with sfermions to the common $m_0$.

ATLAS and CMS experimental searches for SUSY usually present results only in two-dimensional slices of the parameter space of some simplified model assuming fixed values for other parameters [12, 13]. Due to complicated dependence of physical masses and thus experimental signatures on all the model parameters, it is easy to leave specific corners of the model space unexplored in such an approach, leaving out regions where experimental search may have large discovery potential. This has lead several theoretical groups [14, 15] to reinterpret experimental searches in different regions of parameter space with help of simplified simulators of detector response like DELPHES [16] or PGS [17]. This approach can be relatively reliable for moderately simple experimental signatures involving jets and missing transverse energy ($\not{E}_T$), but it cannot be trusted for more difficult experimental objects like photons or tau leptons.

Bayesian parameter inference and MCMC has been successfully employed to find the most viable region of the full parameter spaces, based on requirements that the models should be in accordance with recent experimental constraints [14, 15], including these on the Higgs boson mass and rare $B$-mesons decays. While such scans provide a more complete picture of the still allowed regions of parameter space, few take into account the most relevant experimental signatures to search for or whether these parameter space regions are within experimental reach. This poses difficulties for experimentalists when trying to make direct

use of the results.

In this paper, a general MCMC-based framework for determining the part of non-excluded model parameter space where a given experimental signature can be observed is presented. By adding a signature specific discoverability parameter to the set of current experimental constraints, the interesting regions of the parameter space are found. Models from these regions are then clustered according to phenomenology with use of *g-means* [18] algorithm to enable an automatized construction of reference points for optimizing experimental searches. This step distinguishes our approach from existing similar frameworks for example [19]. The procedure is applicable to a wide range of signatures and models. In order to provide a proof of concept, a concrete example defining a non-excluded part of CMSSM parameter space which could be discoverable with $\tau$-leptons in the 2012 LHC data is outlined.

The paper is structured as follows: section 2 discusses the publicly available software tools used to calculate low energy CMSSM observables, scan and clustering algorithms, as well as the specific constraints and phenomenological parameters used. Section 3 describes the results of the scan and the phenomenological reference points constructed. Appendix 5 explains the details of the algorithm implementation and presents cross-checks of the effects of experimental constraints with other existing results. In Section 4, a summary and comments on the procedure are provided.

## 2 Algorithms and Tools

Experimental constraints on dark matter relic density $\Omega h^2$ as well as on rare processes such as $B_s \to \mu\mu$ and $b \to s + \gamma$ set strict bounds on the parameter space of CMSSM (see for example [15]). Furthermore, if the recent indications of a Higgs boson with mass around 125 GeV [20] turns out to be correct, the fraction of viable models within current experimental reach is extremely small. This renders simple uniform scans highly inefficient. A rough random scan made to explore the parameter dependence in CMSSM, gave a fraction of $10^{-5}$ models in accordance with current experimental constraints. Therefore, more advanced techniques need to be employed to get a representative picture of the discoverable and non-excluded regions of parameter space in an efficent way.

The approach used in this paper is to employ a likelihood distribution $P$, that reflects how well models fit the data, to perform a guided random walk through parameter space using Markov Chain Monte Carlo (MCMC) [21]. This increases the search efficiency as the parameter space is sampled according to the distribution $P$ thus less time is spent sampling low likelihood regions. In this work an adaptive MCMC is implemented, where the likelihood map is based on the compatibility of low energy properties of CMSSM models with experimental and theoretical constraints. These properties are calculated using several

publicly available software tools.

### 2.1 Software Tools

A series of publicly available software tools is used to calculate the low energy parameters needed to check experimental constraints on the SUSY models, and to construct the likelihood map used in the MCMC scan. Parameters are passed between the different tools using the SLHA-interface [22]. The tools are called in sequence starting with the least computationally costly, and after each step the likelihood is updated based on the available parameters. Each component ($i$) of the likelihood is constructed to have a maximal value of $P_i$=1 so that the likelihood always decreases as the chain progresses. This makes it possible to check for rejection after every step in the tool sequence, and enables early termination of the calculations for a large fraction of low likelihood models.

In the first step, `ISAJET` with `isaRED` [23] is used to run the GUT scale universal parameters down to the electroweak scale, calculate $\Gamma(B_s \to \mu\mu)$, and to check weather the models are allowed by several theoretical constraints, including requirements of a $\tilde{\chi}_1^0$ LSP and correct electroweak symmetry. In the next step, `FeynHiggs` [24] and `HiggsBounds` [25] are used to recalculate and check if the model fulfills experimental constraints on the Higgs sector. Afterward, the dark matter relic density, $\Omega h^2$, and $\Gamma(b \to s + \gamma)$ is calculated using `darkSUSY` [26], which also checks against experimental constraints on sparticle masses from LEP $\Delta\rho$ and Z-width (see for example [27, 28]). Finally, 1000 pp signal events at $\sqrt{s} = 8$ TeV are generated using `Pythia` [29] in order to get a leading order estimate of the SUSY cross-section, $\sigma_{\rm LO}$, and to calculate the fraction of events ($\Gamma_\tau, \Gamma_{jet} \ldots$) containing respectively at least one $\tau$, $e$, $\mu$, jet with pseudorapidity in the central part of the detector, $|\eta| < 2.5$, and sufficiently large momentum in the plane perpendicular to the beam axis, $p_T > 20$ GeV, and the average number of these objects per SUSY event ($n_\tau, n_{jet} \ldots$). For each of these objects, the average $p_T$ is calculated for the two ones with the highest transverse momentum. The average missing transverse energy per event, $\not{E}_T$, is also calculated.

The software tools and their employment are summarized in Table 1.

**Table 1**: Software tools and resulting information employed in this work. Average values from `Pythia` are for final states objects with $|\eta| < 2.5$ and $p_T > 20$GeV.

| **Tool** | **Information used** |
|---|---|
| `ISAJET` 7.81 & `isaRED` | SUSY masses, $\Gamma(B_s \to \mu\mu)$ |
| `FeynHiggs` 2.9.2 & `HiggsBounds` 3.7.0 | Higgs sector |
| `darkSUSY` 5.0.5 | $\Omega_\chi h^2$, $\Gamma(b \to s + \gamma)$ |
| `Pythia` 8.162 | $\sigma_{\rm LO}$,$\Gamma$, $\langle n, p_{T1}, p_{T2}\rangle$ for $\tau, e, \mu$, jet,$\langle \not{E}_T\rangle$ |

### 2.2 Likelihood Map and Experimental Constraints

The likelihood map $P$ used to explore CMSSM parameter space is constructed by combining a likelihood $P_{\rm exp}$ based on experimental and theoretical constraints with an ad-hoc likelihood related to the expected number of events with tau leptons, $P_\tau$. Here $P_\tau$ is based on the probability of producing observable $\tau$-leptons with $15/fb$ of the LHC data collected in 2012. $P_\tau$ can be easily replaced by another likelihood function related to observability of any signal of interest. The likelihoods are normalized so that each individual contribution $P_i$ has a maximal value $\max(P_i) = 1$. Thus, the full likelihood becomes:

$$P_{\rm tot} = P_{\rm exp} \cdot P_\tau \quad and \quad P_{\rm exp} = \prod_i P_i \,, \tag{2.1}$$

where $P_i$ are the likelihoods related to experimental limits and theoretical constraints. Some of $P_i$ are either 0 or 1 as specified in table 2. These include most of theoretical constraints, limits checked internally by the software tools used as well as the new experimental limits on $B_s \to \mu\mu$ limits [30], $\Gamma\left(B_s \to \mu\mu\right) < 4.5 \cdot 10^{-9}$. For other experimentally measured quantities Gaussian errors are assumed and the resulting likelihoods are continuous. For example Gaussian distributions around the central experimental values are used for $\Gamma\left(b \to s + \gamma\right)$ and the Higgs mass, while for the relic density a uniform distribution is chosen with a Gaussian tail above the best observational value. The latter accepts models which give the relic density lower than the WMAP result [31], allowing for other unknown sources except of CMSSM neutralinos to contribute to the relic density. The central values and standard deviations used are $\Omega h^2 = 0.1126 \pm 0.0036$ for the relic density [31] and $\Gamma\left(b \to s + \gamma\right) = (3.55 \pm 0.42) \cdot 10^{-4}$ [32] for the charmless $b$-quark decay, where a theoretical uncertainty $\sigma_{th} = \pm 0.33 \cdot 10^{-4}$ is assumed [33]. The central value for the Higgs mass $m_{h0} = (125 \pm 1)$ GeV is assumed in agreement with the recent ATLAS and CMS results. The experimental and theoretical constraints are summarized in table 2. The 2011 and 2012 ATLAS and CMS results of direct searches for SUSY in $R$-parity conserving channels are not included in the present work. The reason for it is two-fold. Firstly, the Higgs mass of $m_{h0} = (125 \pm 1)$ GeV translates in CMSSM into rather high sparticle masses, where the present direct searches results have little effect. Secondly, being members of the ATLAS Collaboration we know how unreliable the translation of present limits into other region of parameter space can be, whenever one uses only approximate modelling of the detector response. We thus prefer to use this opportunity to provide tools to experimenters so that they can chose somewhat more interesting regions of SUSY parameter space to present their results.

The discoverability likelihood $P_\tau$ is constructed as

$$P_\tau = P_K \cdot P_D \,, \tag{2.2}$$

**Table 2**: Experimental and theoretical constraints used and the associated likelihoods

| **Constraints** | **Likelihoods $P_i$** | **Values** |
|---|---|---|
| $\tilde{\chi}_1^0$ LSP, Correct EWSB, No tachyons ... | OK: 1 Not OK: 0 | `ISAJET` 7.81 |
| Sparticle masses, $\Delta\rho$, Z-width | OK: 1 Not OK: 0 | `darkSUSY` 5.0.5 |
| OK Higgs sector | OK: 1 Not OK: 0 | `HiggsBounds` 3.7.0 |
| Branching fraction $B_s \to \mu\mu$ | OK: 1 Not OK: 0 | $\Gamma(B_s \to \mu\mu) < 4.5 \cdot 10^{-9}$ |
| Relic density $\Omega h^2$ | $\exp\left[\frac{(\Omega h^2 - \min(\Omega h^2, \mu_\Omega))^2}{-2\sigma_\Omega^2}\right]$ | $[\mu_\Omega, \sigma_\Omega] = [0.1126, 0.0036]$ |
| Branching fraction $b \to s + \gamma$ | $\exp\left[\frac{(\Gamma_{\rm bsg} - \mu_{\rm bsg})^2}{-2\sigma_{\rm bsg}^2}\right]$ | $[\mu_{\rm bsg}, \sigma_{\rm bsg}] = [3.55, 0.42] \cdot 10^{-4}$ |
| Higgs mass $m_{h0}$ | $\exp\left[\frac{(m_{h0} - \mu_{h0})^2}{-2\sigma_{h0}^2}\right]$ | $[\mu_{h0}, \sigma_{h0}] = [125, 1]$ GeV |

with a discrete part $P_K$ requiring that either the lightest chargino or the second lightest neutralino is heavier than the stau. This is motivated by the fact that in CMSSM $\tau$-leptons are mainly produced through the decay of either $\tilde{\chi}_2^0$ or $\tilde{\chi}_1^\pm$. Continuous likelihood $P_D$ is a Poisson *discoverability* measure constructed from the sum of likelihoods for observing a given number of tau events, $N_\tau \geq 1$, given the expected number of events containing at least one $\tau$, $\langle N_\tau \rangle = \Gamma_\tau \cdot \mathcal{L} \cdot \sigma_{LO}$.

$$P_D = \sum_{N_\tau = N_\tau^{\min}} P(N_\tau | \langle N_\tau \rangle) \quad , \quad P(N_\tau | \langle N_\tau \rangle) = \frac{\langle N_\tau \rangle^{N_\tau} \exp\left[ - \langle N_\tau \rangle \right]}{N_\tau} \, . \tag{2.3}$$

Leading order SUSY cross-section $\sigma_{LO}$ in $pp$ collisions at 8 TeV center-of-mass energy, luminosity of $15/fb$ and the fraction of events containing at least one $\tau$,$\Gamma_\tau$, as found from `Pythia`, are used to calculate $\langle N_\tau \rangle$ above.

The search range in CMSSM parameter space follows the suggestion in [34]. The ranges for $m_0$, $m_{1/2}$, $A_0$ and $\tan\beta$ are presented in table 3. Only positive values of sign($\mu$) are

considered where the muon anomalous magnetic moment, $(g-2)_\mu$, constraint seem to be easier to satisfy [15], but this constraint is not included in the present work. A fixed value for $m_{\rm top} = 173$ GeV was taken.

The lower bounds on the universal masses and on $\tan\beta$ in table 3 stem from LEP [28, 35] bounds, while the upper limits are chosen on the basis of naturalness for the masses $m_0$ and $m_{1/2}$ and perturbativity of the Yukawa couplings for $\tan\beta$. The range for the trilinear coupling $A_0$ is extended compared to [34], since the Higgs mass has a quadratic dependence on $A_0$ [36] allowing for higher $m_{h0}$ at large values of $|A_0|$.

It is important to note that the more realistic the likelihood, the more computationally

**Table 3**: Search ranges used for the CMSSM MCMC scans.

| Parameter | Range |
|---|---|
| $m_0$ | [60,3000] GeV |
| $m_{1/2}$ | [60,3000] GeV |
| $A_0$ | [-5000,5000] GeV |
| $\tan\beta$ | [2,60] |
| sign($\mu$) | +1 |

efficient is the search for interesting models. However, the set of models found does not depend on fine details of the likelihood function used, as we accept points in a range of likelihood. Our goal is to find a set of interesting models fulfilling latest experimental constraints, in order to guide further experimental searches. We do not intend to make any *statistically quantified decision* on which of the selected models are more likely than others.

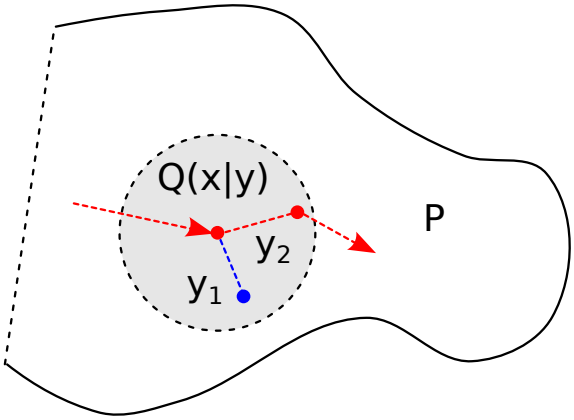
### 2.3 MCMC Algorithm

The MCMC method used here is a Metropolis-Hastings algorithm [37] where, given a point $\mathbf{x} = \{x_1, x_2 \dots x_D\}$ in a D-dimensional parameter space, a proposal distribution, $Q(\mathbf{y}|\mathbf{x})$, is used to sample a new point $\mathbf{y}$. The proposal distribution is related to the likelihood $P$ of the new point $\mathbf{y}$ being "interesting" from the point of view of requirements described in section 2.2. The new point is accepted randomly with a probability given by

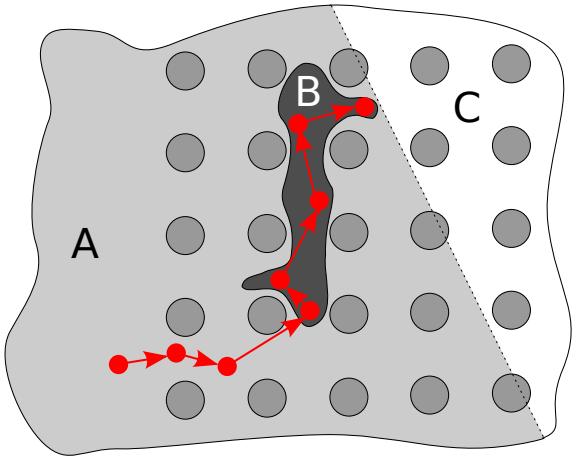
$$\alpha(\mathbf{y}|\mathbf{x}) = \min\left(1, \frac{P(\mathbf{y})Q(\mathbf{x}|\mathbf{y})}{P(\mathbf{x})Q(\mathbf{y}|\mathbf{x})}\right) , \tag{2.4}$$

If $\mathbf{y}$ is accepted, it is added to the chain and the next point is sampled starting from $\mathbf{y}$. If it is not accepted, the chain remains at $\mathbf{x}$ and the process is repeated as illustrated in figure 1. The asymptotic distribution of likelihoods calculated for the resulting chain of points is the desired likelihood distribution $P$.

In order to efficiently map possible high likelihood regions of the parameter space which are separated by large regions of low likelihood a regional adaptive MCMC algorithm similar to [38] has been implemented. The algorithm approximates the target likelihood distribution $P$ as a mixture of normalized multivariate Gaussian distributions and uses this approximation as a basis for a proposal $Q(y|x)$, as explained in the following section 2.3.1. This proposal is used to guide multiple MCMC search chains in parallel and it is iteratively updated according to the resulting selected sample of points.




a) A proposal distribution $Q(\mathbf{y}|\mathbf{x})$ used to sample new points $(\mathbf{y_1}, \mathbf{y_2})$ from a point $\mathbf{x}$. These points are either accepted $(\mathbf{y_2})$ or rejected $(\mathbf{y_1})$ depending on the ratio between the underlying likelihood $P$ and $Q$ as given by $\alpha$ (2.4)




b) Example MCMC random walk seeking out narrow high likelihood regions $A$, from an initial starting point in a low $(A)$ or zero $(C)$ likelihood region. Th grid of dots is shown to illustrate how such regions can be missed by uniform grid-based scans.

**Figure 1**: Illustrations of the standard MCMC sampling method, 1a, and a typical MCMC random walk, 1b

### 2.3.1 Adaptive Multi Chain Monte Carlo

An initial estimate for the proposal is constructed by uniformly sampling the space such that all separated regions where the likelihood $P$ is high are covered. The points of CMSSM parameter space chosen for the initial sample are required to pass all discrete cuts and to give experimentally measured physical variables within a reasonable range of the experimentally preferred values, see table 2 for details. Sampled points in CMSSM parameter space are weighted according to their likelihood and clustered using *k-means* algorithm, see section 2.4. For each cluster of points we estimate the shape in CMSSM parameter space by calculating the the weighted mean $\boldsymbol{\mu}$ vector and and covariance matrix $\boldsymbol{\Sigma}$

The number of clusters corresponds to the number of normalized Gaussian distributions (normal mixture) that is to be used to approximate the likelihood distribution $P$, as explained below.

Efficient sampling of high likelihood regions in the parameter space separated by large regions of low likelihood required a modification of the proposal in order to introduce a small fraction of large jumps [38, 39]. To this end a global proposal term $q_G(\mathbf{y})$, has been introduced in addition to the standard local one, $q_L(\mathbf{y}|\mathbf{x})$. The summary proposal distribution becomes:

$$Q(\mathbf{y}|\mathbf{x}) = \beta q_L(\mathbf{y}|\mathbf{x}) + (1-\beta) q_G(\mathbf{y})\,, \tag{2.5}$$

Here $\beta$ is a mixing parameter relating the global and the local proposal terms, explained further.

The global proposal $q_G(\mathbf{y})$ is taken as the weighted mixture of a set of $m$ multivariate normal distributions, $\mathcal{N}$, as in [40]. The number of multivariate normal distributions used is chosen high enough to describe the main features of the target distriubtion (number of seperated regions, curvature, etc), while each weight factor $w_i$ in the formula below is estimated by summing up the total likelihood over CMSSM parameter space points in a cluster $i$.

$$q_G(\mathbf{y}) = \sum_{i=1}^{m} w_i \mathcal{N}(\mathbf{y}|\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_{G,i}) \quad , \quad \mathcal{N}(\mathbf{y}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{\exp\left[-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\boldsymbol{\mu})\right]}{\sqrt{(2\pi)^D\,|\boldsymbol{\Sigma}|}}\,, \tag{2.6}$$

Here $\boldsymbol{\mu}_i$ is the vector of the means of the $i$'th normal distribution, while $\boldsymbol{\Sigma}_{G,i}$ is the covariance matrix of the $i$'th component of the mixture. The local proposal $q_L(\mathbf{y}|\mathbf{x})$ is taken as a normal distribution with mean, $\mathbf{x}$, and the covariance, $\boldsymbol{\Sigma}_{L,i}$, characterizing the closest cluster in the parameter space. An euclidean distance measure is used and each parameter is scaled so that the search ranges defined in table 3 vary from 0 to 1. The local proposal covariance is chosen so that: $\boldsymbol{\Sigma}_{L,i} = \alpha_i \boldsymbol{\Sigma}_{G,i}$ , $i \in \{1, 2, \ldots m\}$, where $\alpha_i$ is a parameter adapted such that the local acceptance rate for points in the parameter space region within the cluster $i$ is between 0.05 and 0.15. The rather low acceptance rate is chosen because the hierarchical nature of the likelihood calculation yields higher computational speed for low acceptance rates. Thus our optimal acceptance rate is probably lower than that of 0.23 found in [41]. The acceptance probability for stepping from a given point $x$ to a new point $y$ is then given as:

$$\alpha(\mathbf{y}|\mathbf{x}) = \min\left(1, \frac{P(\mathbf{y})\left[\beta q_L(\mathbf{x}|\mathbf{y}) + (1-\beta) q_G(\mathbf{x})\right]}{P(\mathbf{x})\left[\beta q_L(\mathbf{y}|\mathbf{x}) + (1-\beta) q_G(\mathbf{y})\right]}\right)\,. \tag{2.7}$$

The search chains are started from random CMSSM parameter space points in the weighted sample and followed independently. After a given number of steps data is re-clustered and the proposals are updated, taking into account the new sampled parameter space points, where the new points are weighted according to the estimated likelihood. A fraction of points in each chain is discarded to ensure sampling from the equilibrium part of $P$ distribution. This is done by requiring a certain likelihood threshold $P_{\min}$ for the first relevant point in each chain [21, Ch. 1.11.4]. The implementation details are described in the Appendix 5.

### 2.4 Clustering Algorithm

A suitable clustering algorithm based on *k-means* [42] has been devised in order to cluster likelihood-weighted points in CMSSM parameter space. The role of clustering is two-fold. Firstly clusters are needed to calculate the approximate Gaussian distributions used in the proposal described in the previous section 2.3.1. Secondly, after the choice of the set of high-likelihood model-points in CMSSM parameter space is made, these points are clustered according to the different experimental signatures they are expected to exhibit in the detectors at the LHC.

The *k-means* algorithm defines clusters in the parameter space by assigning each space point to the closest centroid, $(C)$. The algorithm is initialized by choosing at random $k$ points in the parameter space as cluster centers $C$. Next, each $C_i$ is refined as the average of the points near to it and points are reassigned to the new $C$, and the procedure is repeated until it converges to a set of stable $C$s.

A number of improvement to the standard *k-means* [42] procedure was employed in this article in order to cure several undesirable features: arbitrary choice of the number of clusters, $k$, slow convergence and sensitivity to the "outlying" parameter space points. To increase the speed of the algorithm, a maximum number of iterations and a minimum improvement between iterations was set for the centroids positions refinement. Furthermore the speed to find the nearest neighboring points was improved using a multidimensional binary search tree (*kd*-tree) as in [43] resulting in the parameter space being searched for the points closest to a given centroid along alternating dimensions defined by the *kd*-tree branches. The speed was further gained by grouping these branches such as to minimize the close points population [44].

The random first guess of centroids positions was improved as suggested in [45], (*k*-means++ algorithm). Here the probability of picking a new point was weighted by the square distance to the closest already picked point. This guess reduced as well the average number of iterations to achieve convergence.

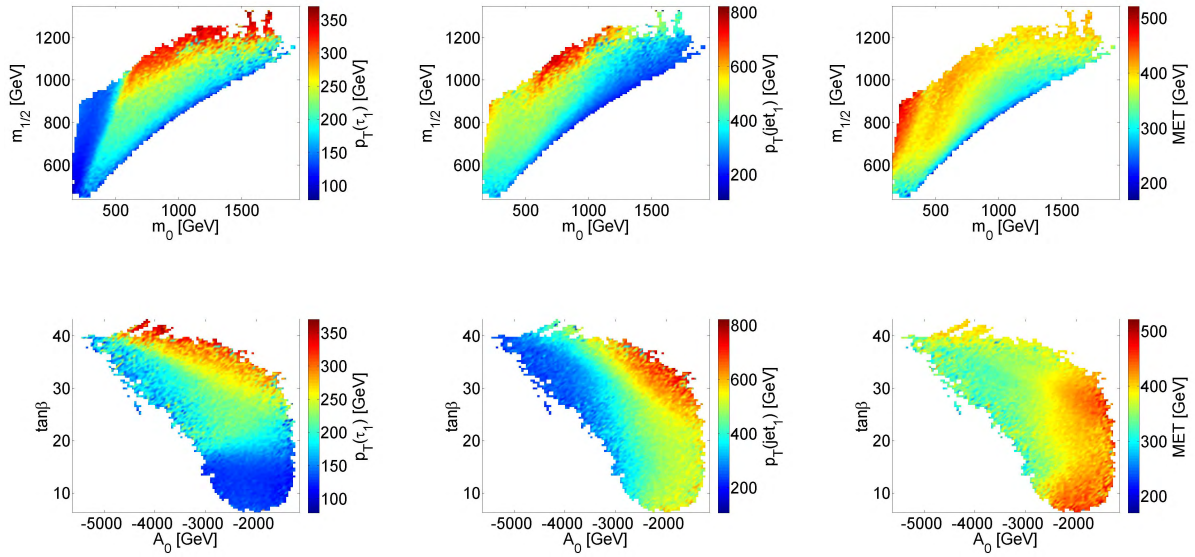
To determine the number of clusters, $k$, *g-means*-algorithm [18] was employed. It started with applying *k-means* for $k = 2$, thus dividing the parameter space into two sub-clusters. Then a statistical test to verify if the likelihood distributions of both clusters can be described by a single Gaussian was performed. If the test rejected the single Gaussian distribution hypothesis, the procedure was repeated recursively for each of the new clusters, otherwise recursion was terminated. The statistical test was done by means of the dimensional Anderson-Darling normality measure [46]. The distances between the points and a plane that separated clusters and perpendicular to the vector between the two centroids were subjected to the measure above. As the clustering could be sensitive both to outliers and to the random positioning of initial centroids, the whole splitting procedure

was iterated $n_{\text{avg}}$ times and the average number of clusters $\langle k \rangle$ was noted. Next, the cluster with $k$ closest to $\langle k \rangle$ was picked. If there were several equidistant clusters, one of these was picked at random. In the final step the obtained centroids were subjected to the *k-means* clustering one more time to ensure that a stable configuration has been found. The *k-means* was repeated a very large number of times in the clustering procedure. Thus the speed optimizations described above played an important role in ensuring the algorithm was usable for this work.

## 3 CMSSM with $\tau$ Signatures

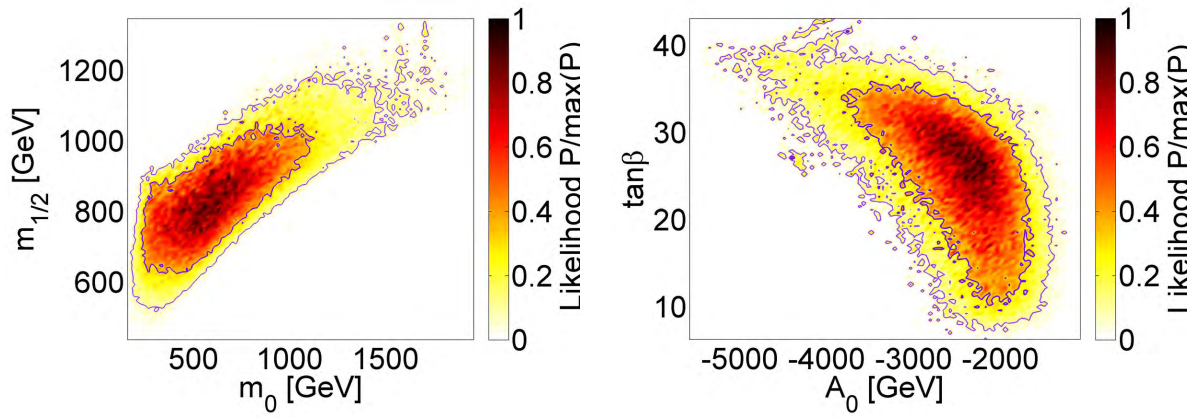
### 3.1 Results

All viable models found have large negative values of $A_0$ in common, but otherwise span a relatively large range of sparticle masses and values of $\tan\beta$. The ranges for the mean $\not{E}_T$ and $p_T$ for leading jet and $\tau$ are shown in the CMSSM mass planes $m_0 - m_{1/2}$ and $A_0 - \tan\beta$ in figure 2 and two dimensional likelihood distributions are shown in figure 3. The distributions are constructed by binning the models into $N_{\text{bins}} = 100$ bins along each dimension, where the likelihood of each bin is approximated by the number of models contained. The effects of different constraints separately are discussed in the Appendix 5.




$p_T(\tau_1)$ $p_T(jet_1)$ $\not{E}_T$

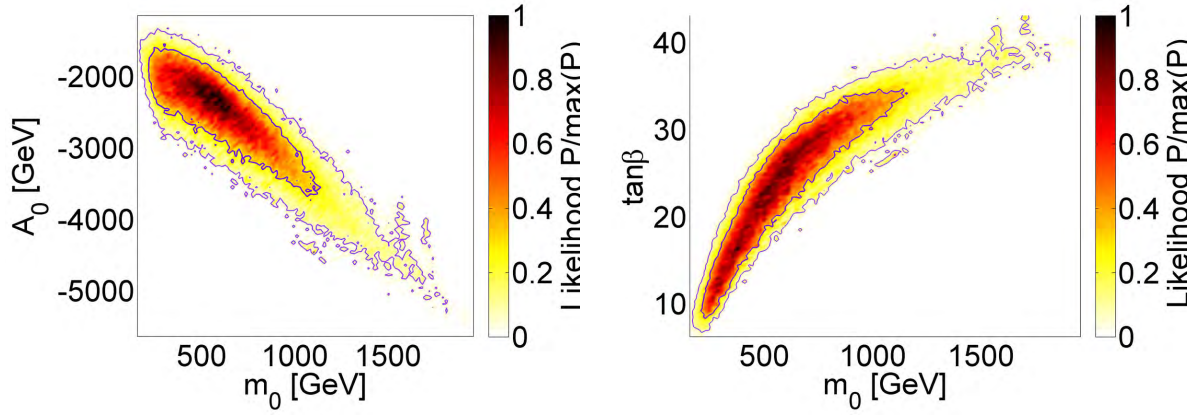
**Figure 2**: Average value per bin for mean $\not{E}_T$ and $p_T$ for leading jet and $\tau$ shown in the CMSSM mass plane $m_0 - m_{1/2}$ (above) and $A_0 - \tan\beta$ plane (below)




$m_0$ vs $m_{1/2}$ plane.



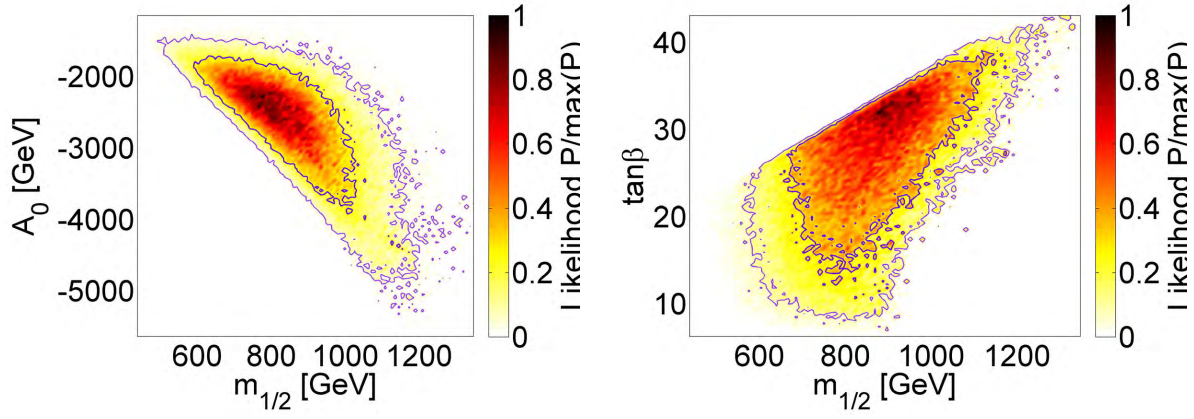

$A_0$ vs $\tan\beta$ plane.




$m_0$ vs $A_0$ plane.




$m_0$ vs $\tan\beta$ plane.




$m_0$ vs $A_0$ plane.




$m_{1/2}$ vs $\tan\beta$ plane.

**Figure 3**: Marginalized likelihood maps for different planes in CMSSM space.

The *discoverability* likelihood constrains the SUSY $\tau$ production cross-section to not be too small. This sets upper bounds on how large the gaugino and scalar masses can be since the production cross-section falls sharply as the masses of colored sparticles grow. The cross-section has as well a slight $A_0$ dependence which allows for higher masses at higher negative values of $A_0$.

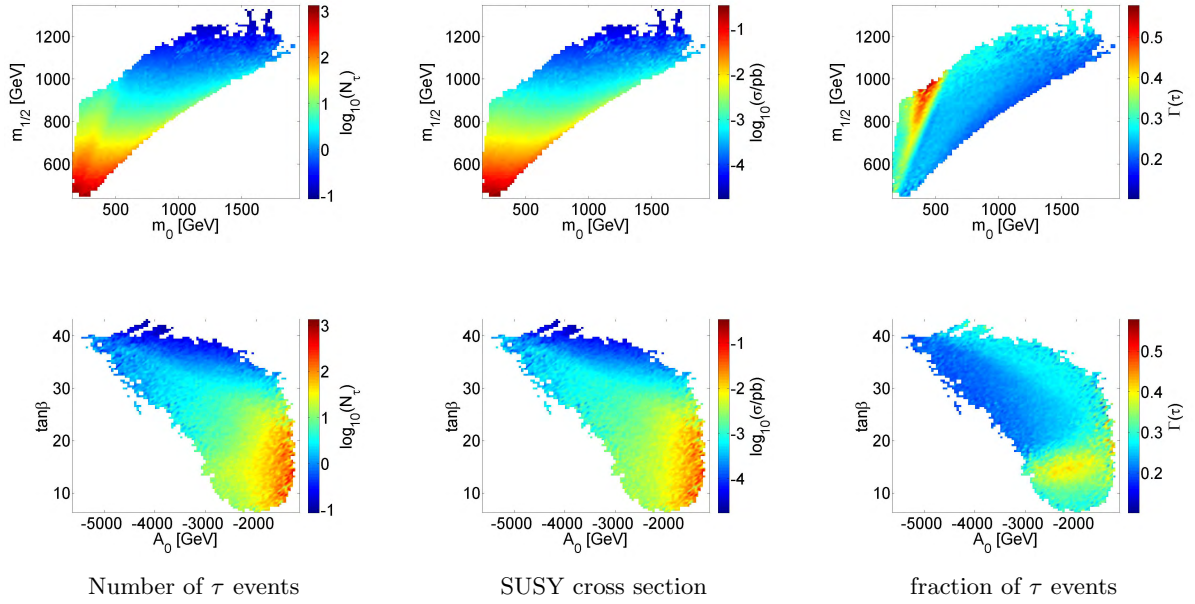



**Figure 4**: Average value per bin for number of $\tau$ events, SUSY cross-section, and the fraction of $\tau$ events, in the mass plane $m_0 - m_{1/2}$ (above) and $A_0 - \tan\beta$ (below)

### 3.2 Phenomenology and Reference Points

The relatively wide range of values for SUSY masses and values of $\tan\beta$ found leads to a wide range of values of phenomenological properties such as average $\not{E}_T$,the average missing energy per SUSY event, $p_T(\tau_1), p_T(\text{jet}_1)$, the average $p_T$ of the leading $\tau$ and the leading jet, and $n_\tau, n_{\text{jet}}$, the average number of $\tau$'s/jets per SUSY event, see 2.

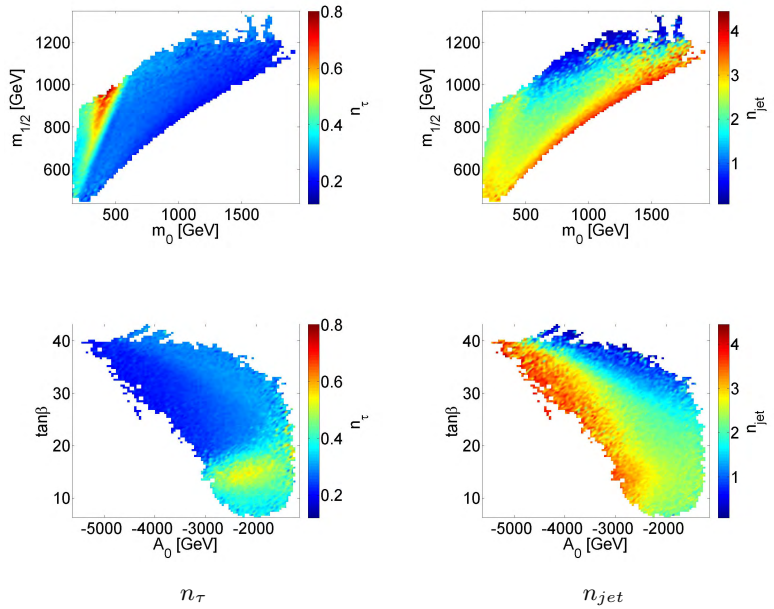
**Table 4**: Range and best fit value for phenomenological parameters.

| $\not{E}_T$ [GeV] | | | $n_{\text{jet}}$ | | | $\text{jet}_1(p_T)$ [GeV] | | | $n_\tau$ | | | $\tau_1(p_T)$ [GeV] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| min | cent | max | min | cent | max | min | cent | max | min | cent | max | min | cent | max |
| 1 | **386** | 523 | 0.0 | **2.5** | 4.5 | 55 | **466** | 857 | 0.0 | **0.6** | 2.0 | 55 | **151** | 372 |

The $p_T$ values for the leading jet and $\tau$ lepton obviously tend to be higher for high sparticle masses, since higher masses in CMSSM lead to higher mass splittings between the sfermions

and the LSP. The $p_T$s also become larger with $A_0$ closer to 0 and for high $\tan\beta$ values. The missing energy on the other hand tends to become larger at smaller scalar masses and increasing gaugino mass. The increase in $\not{E}_T$ with $m_{1/2}$ is likely due to increasing neutralino mass. These dependencies are illustrated in figure 2.

The average number of $\tau$s per SUSY events is mostly due to the branching fraction in to $\tau$s as it can be seen comparing figure 5 and 4. One tau with high $p_T$ per event is produced on average. The SUSY branching fraction to $\tau$s is largest at low values of $\tan\beta$ and $m_0$. At least one high $p_t$ jet is expected in almost every event. The average numbers of jets increases with $m_0$, $\tan\beta$ and $|A_0|$.




$n_\tau$ $n_{jet}$

**Figure 5**: Average value per bin for the mean number of $\tau$s and jets per SUSY event shown in the $m_0 - m_{1/2}$ plane (above) and $A_0 - \tan\beta$ plane (below)

In order to construct reference benchmark models that cover these different phenomenological properties the sample was clustered according to the phenomenological observables: $\{\not{E}_T, n_{\text{jet}}, p_T(\text{jet}_1), n_\tau, p_T(\tau_1)\}$,
details are described in the Appendix 5.

With these, nine phenomenological clusters shown in table 5 were found. The SUSY and model related parameters for these clusters are shown in table 6. The centroids of the clusters can be regarded as reference (benchmark) points. As it can be seen in table 5, around 160 events with high energetic taus in the central part of the detector are expected for the highest cross-section reference point, this is for $15/fb$ of 8 TeV CMS energy at the LHC. Even if predicted 2012 luminosity per experiment is around $25/fb$, this number of

events might not be enough to detect the signal. Exploration of the other reference points might have to wait until the LHC 13 TeV operations.

Various graphical projections of the clusters are shown in figures 6 and 7. It is clear from figures 6 that one finds viable models lying in the tails of clusters, far away from centroid positions. These models exhibit either low jet activity in the central part of the detector, but produce high $p_t$ tau leptons, or have low number of high $p_t$ jets (monojets) and low momentum taus. We have not investigated these models further yet, but it is clear that standard LHC SUSY searches assuming presence of high $p_t$ jets for triggering purpose might fail for such models.

Table 8 gives branching fractions of the lightest Higgs boson for the reference points, compared to the SM Higgs ones. Alas, the branching fractions do not differ by more than 5% from SM values for these models.

**Table 5**: Phenomenological parameters of clusters found. The first two columns are the cluster index, id, (matches id in table 6) and number of model-points in the cluster, n. For each parameter the cluster centroid value, cent, is listed along with the minimum, min, and maximum, max, for the cluster. Centroid values can be regarded as reference values characterizing given experimental phenomenology.

| id | n | $\not{E}_T$ [GeV] | | | $n_{\text{jet}}$ | | | $\text{jet}_1(p_\text{T})$ [GeV] | | | $n_\tau$ | | | $\tau_1(p_\text{T})$ [GeV] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | min | cent | max | min | cent | max | min | cent | max | min | cent | max | min | cent | max |
| 1 | 17502 | 61.0 | **384.9** | 467.2 | 0.8 | **2.8** | 3.9 | 159.9 | **424.9** | 581.1 | 0.3 | **0.4** | 1.1 | 60.7 | **129.7** | 218.7 |
| 2 | 22771 | 3.7 | **438.3** | 523.0 | 0.1 | **2.3** | 3.8 | 149.2 | **535.8** | 698.4 | 0.4 | **0.5** | 2.0 | 57.3 | **137.7** | 345.6 |
| 3 | 33412 | 362.9 | **432.5** | 516.4 | 1.0 | **1.8** | 2.7 | 451.6 | **582.1** | 730.4 | 0.2 | **0.3** | 0.5 | 105.9 | **242.0** | 308.0 |
| 4 | 39476 | 266.1 | **332.5** | 396.2 | 2.3 | **3.1** | 4.1 | 219.4 | **288.6** | 418.7 | 0.1 | **0.2** | 0.4 | 61.6 | **169.8** | 240.2 |
| 5 | 21767 | 0.6 | **274.7** | 339.2 | 0.8 | **3.6** | 4.5 | 130.6 | **232.2** | 362.9 | 0.1 | **0.2** | 0.8 | 55.2 | **137.7** | 207.7 |
| 6 | 26284 | 293.4 | **398.5** | 494.7 | 1.1 | **2.6** | 3.6 | 278.8 | **372.5** | 594.3 | 0.0 | **0.2** | 0.3 | 56.0 | **138.0** | 225.7 |
| 7 | 40229 | 172.1 | **350.9** | 432.9 | 0.0 | **2.4** | 3.1 | 55.4 | **391.6** | 519.9 | 0.1 | **0.3** | 0.6 | 149.5 | **218.5** | 336.8 |
| 8 | 43078 | 225.2 | **389.2** | 449.0 | 0.2 | **1.9** | 2.7 | 218.0 | **477.2** | 619.1 | 0.1 | **0.3** | 0.5 | 168.9 | **250.6** | 331.6 |
| 9 | 39616 | 186.6 | **395.1** | 467.8 | 0.0 | **1.1** | 1.8 | 315.4 | **633.8** | 856.6 | 0.1 | **0.3** | 0.7 | 213.3 | **291.2** | 371.9 |

## 4 Conclusions

This work presents a new method for finding and classifying SUSY models that can be potentially discovered in an accelerator experiment, here LHC experiments. The method uses an adaptive MCMC algorithm to find interesting models and uses a clustering algorithm to classify the models according to phenomenology. The likelihood map is constructed using an extendible tool chain that incorporates recent limits from multiple sources through the SLHA interface. As an example and test of the method the highly constrained CMSSM

**Table 6**: CMSSM parameters, relict density, 8 TeV CMS LHC production cross-section and the number of expected events with $\tau$ leptons for the centroids of clusters. All models have sign$\mu > 0$ and $m_{\rm top} = 173$ GeV.

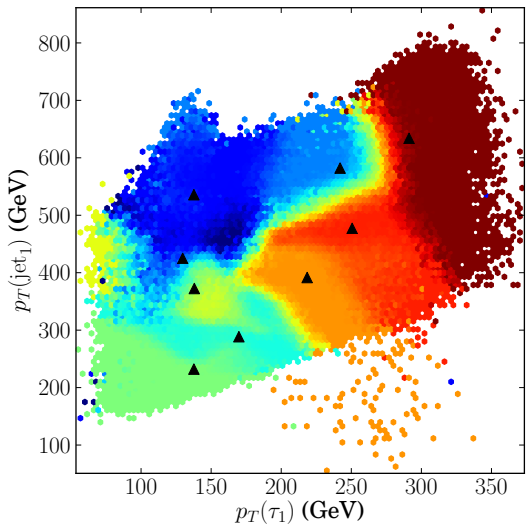
| **id** | $m_0$ [GeV] | $m_{1/2}$ [GeV] | $A_0$ [GeV] | $\tan\beta$ | $\langle N_\tau \rangle$ | $\ln P$ | $\Omega h^2$ | $\sigma$ [fb] |
|---|---|---|---|---|---|---|---|---|
| 1 | 284.0 | 614.4 | -1832.4 | 14.5 | 162.9 | -2.2 | 0.1 | 30.2 |
| 2 | 311.8 | 795.7 | -2090.1 | 13.2 | 19.5 | -1.2 | 0.1 | 3.1 |
| 3 | 615.4 | 793.4 | -2340.3 | 26.4 | 11.1 | -0.9 | 0.1 | 2.6 |
| 4 | 973.3 | 919.7 | -3466.2 | 30.2 | 4.8 | -0.9 | 0.1 | 1.5 |
| 5 | 726.8 | 915.2 | -3363.2 | 22.9 | 9.9 | -1.6 | 0.1 | 3.1 |
| 6 | 658.6 | 812.2 | -2295.0 | 28.7 | 3.4 | -1.1 | 0.1 | 1.2 |
| 7 | 759.3 | 833.7 | -2837.6 | 28.0 | 7.8 | -0.9 | 0.1 | 2.2 |
| 8 | 841.7 | 903.3 | -2929.3 | 29.9 | 3.3 | -0.8 | 0.1 | 0.8 |
| 9 | 789.7 | 983.9 | -2763.2 | 28.7 | 1.4 | -0.9 | 0.1 | 0.3 |

**Table 7**: Higgs and sparticles masses for the centroids of clusters.

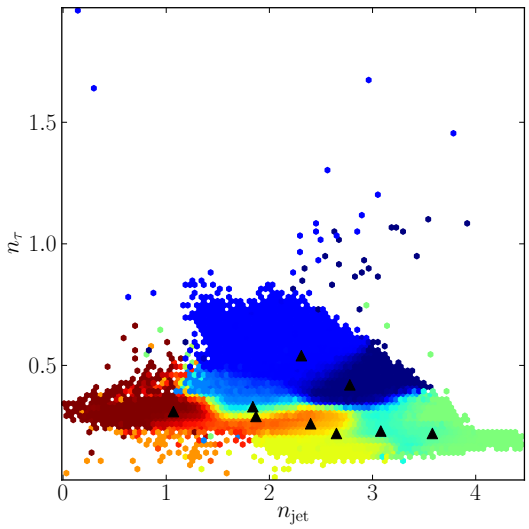
| **id** | $m_{h_0}$ [GeV] | $m_{\tilde{t}_1}$ [GeV] | $m_{\tilde{g}}$ [GeV] | $m_{\chi_1^0}$ [GeV] | $m_{\tilde{\tau}_1}$ [GeV] |
|---|---|---|---|---|---|
| 1 | 123.3 | 967.1 | 2105 | 408.2 | 412.8 |
| 2 | 123.8 | 1016 | 2271 | 444.5 | 447.9 |
| 3 | 124.4 | 844.9 | 1923 | 368.8 | 372.2 |
| 4 | 124.9 | 676.9 | 1632 | 307.9 | 308.1 |
| 5 | 123.9 | 1124 | 1991 | 384.8 | 386.7 |
| 6 | 124.2 | 896.1 | 1975 | 381.9 | 382.7 |
| 7 | 124.9 | 668 | 1421 | 262.8 | 264.3 |
| 8 | 125.1 | 1201 | 2366 | 466.7 | 467.8 |
| 9 | 124.8 | 879.9 | 1867 | 357.5 | 360.4 |

**Table 8**: The lightest Higgs branching fractions in relation to the SM value for the centroids of clusters.

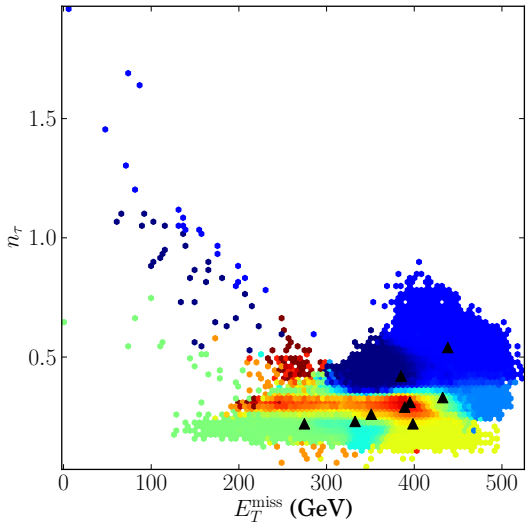
| **id** | $\Gamma^{SUSY}_{H\to ZZ}/\Gamma^{SM}_{H\to ZZ}$ | $\Gamma^{SUSY}_{H\to WW}/\Gamma^{SM}_{H\to WW}$ | $\Gamma^{SUSY}_{H\to\gamma\gamma}/\Gamma^{SM}_{H\to\gamma\gamma}$ | $\Gamma^{SUSY}_{H\to\tau\tau}/\Gamma^{SM}_{H\to\tau\tau}$ | $\Gamma^{SUSY}_{H\to\mu\mu}/\Gamma^{SM}_{H\to\mu\mu}$ |
|---|---|---|---|---|---|
| 1 | 1.03 | 1.04 | 1.02 | 1.04 | 1.04 |
| 2 | 1.03 | 1.05 | 1.03 | 1.03 | 1.03 |
| 3 | 1.03 | 1.05 | 1.03 | 1.03 | 1.03 |
| 4 | 1.03 | 1.04 | 1.03 | 1.04 | 1.04 |
| 5 | 1.03 | 1.05 | 1.03 | 1.04 | 1.04 |
| 6 | 1.04 | 1.05 | 1.03 | 1.03 | 1.03 |
| 7 | 1.03 | 1.04 | 1.03 | 1.04 | 1.04 |
| 8 | 1.04 | 1.05 | 1.03 | 1.03 | 1.03 |
| 9 | 1.04 | 1.05 | 1.03 | 1.03 | 1.03 |



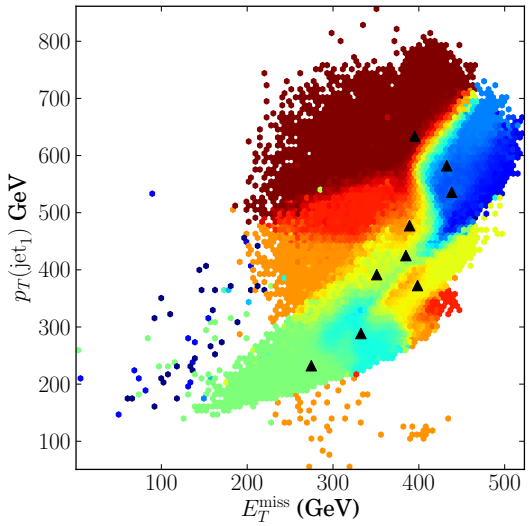

Leading $\tau$ vs jet $p_T$




$n_{\text{jet}}$ vs $n_\tau$ projection.




$\not{E}_T$ vs $n_\tau$ projection.

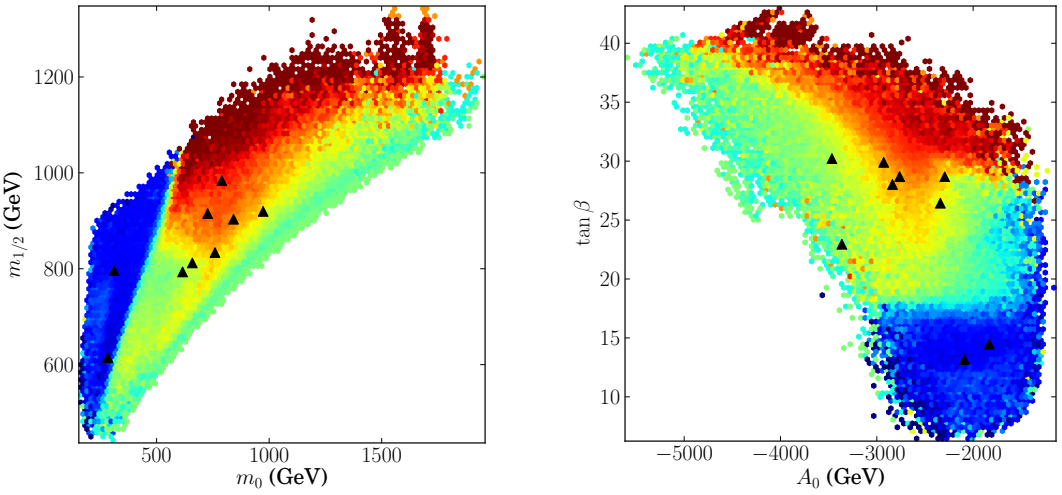



$\not{E}_T$ vs leading jet $p_T$ projection.

**Figure 6**: Projections of clusters. Colors indicate to which cluster a model belongs. Black triangles indicate locations clusters centroids.

parameter space has been searched for models that could potentially be discovered using 2012 LHC with $\tau$-lepton based signatures.

Although simplified models like CMSSM are severely constrained, we are still able to find regions fulfilling recent LHC bounds on Higgs mass and rare $B$-meson decays, and giving relic density in agreement with WMAP results. Nine reference (bench-mark) points




$m_0$ vs $m_{1/2}$ projection.

$A_0$ vs $\tan\beta$ projection.

**Figure 7**: Projections of clusters. Colors indicate to which cluster a given model belongs.

exhibiting different phenomenologies were found with use of a *g-means* clustering algorithm. These reference points can be used to optimize searches. This makes the method very attractive from an experimental point of view, and applying the method to other models and different signatures would be a natural extension of this work.

The method has proven successful in finding and classifying SUSY models, but could still benefit from several extensions and improvements. More constraints could be added to the likelihoods and more advanced statistical analysis of the simulated data could be incorporated. Another interesting prospect would be to include detector simulations using PGS [17] or DELPHES [16] to get somewhat more realistic estimates for the expected signal, although we are skeptical about realism of such simulations. In addition, the MCMC algorithm constructed could still benefit from improvements to increase stability and efficiency, in addition to rigorous numerical testing and proofs of ergodicity. Finally a better distance measure for the clustering could allow for precise predictions of expected discovery potential.

On the more experimental side we find some viable models lying in the tails of clusters formed based of phenomenological observables. These models exhibit either low jet activity in the central part of the detector, but produce high $p_t$ tau leptons, or have low number of high $p_t$ jets (monojets) and low momentum taus. We have not investigated these models further yet, but it is clear that standard LHC SUSY searches assuming presence of high $p_t$ jets for triggering purpose might fail for such models.

## Acknowledgments

This work has been performed in the scope of the Centre for Dark Matter Research (DAMARA) at the Department of Physics and Technology, University of Bergen, Norway. It was funded by the Bergen Research Foundation and the University of Bergen, as well as the Norwegian Research Council in the framework of High Energy Particle Physics project.

We would like to give special thanks Therese Sjursen for fruitful discussions leading to the start-up of this work. We would also like to thank members of the group for Subatomic Physics at University of Bergen for their support.

## 5 Appendix: Scan Implementation, clustering and cross-checks

The implementation of the scan was written in `Python` . It was capable of running multiple MCMC chains in parallel. The scan was initiated using a random sample containing roughly 100 points to find an estimate for the proposal distribution. The initial points were required to pass all discrete constraints, have parameter values close to the central observational values for $\Gamma(b \to s + \gamma)$, $m_{h0}$ and $\Omega h^2$, while having an expected number of produced $\tau$ leptons, $\langle N_\tau \rangle > 1$. The chosen values are shown in figure 9.

**Table 9**: Experimental constraints used in the initial sampling.

| **Constraints** | **Range** |
|---|---|
| $\Omega h^2$ | $[0, 0.2]$ |
| $\Gamma\left(b \to s + \gamma\right)$ | $[3, 4] \cdot 10^{-4}$ |
| $m_{h0}$ | $[122, 128]$ GeV |
| $\langle N_\tau \rangle$ | $[1, \infty)$ |

Five clusters were established from the initial parameter space points with the *k-means* algorithm. This number was found to be sufficient to give a reasonable approximation of the likelihood distribution of the sample. From the initial sample, ten chains were initiated with two chains starting from each cluster. Before sampling started, each chain was required to reach a minimum likelihood to be included in the sample. This was chosen to be $2\sigma$ away from the central value for $\Gamma(b \to s + \gamma)$, $m_{h0}$, $\Omega h^2$ and $\langle N_\tau \rangle = 1$ corresponding to the likelihood, $\ln P_{\min} \sim -3 \cdot \frac{4}{2} - 0.5 = -6.5$.

The proposal distribution was updated at intervals $\Delta N = 10000$ steps by adding the new sample points and recalculating cluster means and covariances. The new points were added

without weights since they were already a product of weighted sampling. For practical purposes an upper size on the clustering sample was set to be 10000 unique points, which was found to be sufficient to give a good proposal estimate. The search chains were run in parallel on ten cores for roughly 500 hours, resulting in a sample size of $N = 2896332$, corresponding to 284496 unique models. From this sample 3938 outliers (corresponding to 361 unique models) with log-likelihood $ln\ P < -6.5$ were removed.

In order to illustrate the effects of the different experimental and theoretical constraints, low energy properties were calculated for 300 000 models sampled uniformly within the search range. The computationally expensive `Pythia` simulations where not done for these models and a looser relic density constraint compared to the one used for MCMC initialization was used to get sufficient data to describe the qualitative features of the constraint.

As a cross-check with the vast literature on the subject (see for example [34, 47, 48]) we briefly describe the effects of the most important constraints by visualizing how the initial selections affect the model density. The effects are illustrated in figure 8. We observe, in agreement with results of [15, 47] that:

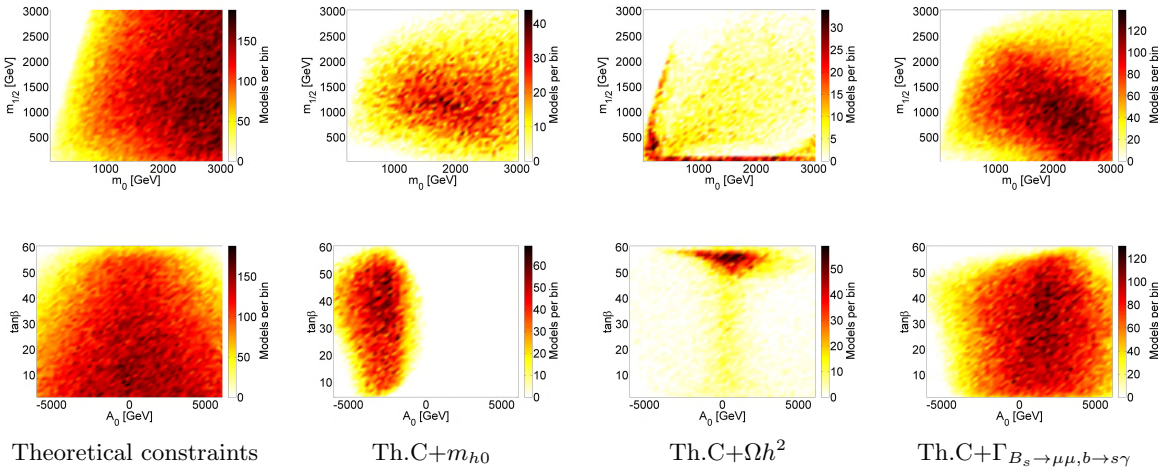
- **Theoretical constraints**
  Theoretical constraints remove the low $m_0$ and $m_{1/2}$ regions primarily avoiding a $\tilde{\tau}_1$-LSP and tachyonic sparticles. The excluded regions becomes larger at large $\tan\beta$ and $|A_0|$, and for large values of $A_0$ a considerable part of the low mass regions, $m_0, m_{1/2} \lesssim 1000$ GeV gives tachyons.

- **Higgs mass $\mathbf{m_{h0} \in [122, 128]}$**
  Requiring a 125 GeV Higgs mass, with positive $\mu$ excludes all positive values of $A_0$ within the selected $m_0, m_{1/2}$-range. For large negative values of $A_0$ however, the $\tilde{t}$-loop corrections to the Higgs mass become large. This is the main reason for the asymmetry in $A_0$ seen in 3. At large values of $m_{1/2}$ and $m_0$ the `FeynHiggs` calculations of the Higgs mass corrections become inaccurate and thus we excluded these regions.

- **Relic Density $\mathbf{\Omega h^2 < 1}$**
  As is well known, the relic density Dark Matter in CMSSM is generally orders of magnitude larger than allowed by WMAP and PLANCK results [48–50], apart from special regions where the relic density is suppressed by resonant neutralino annihilation or co-annihilation cross-sections. The low $m_{1/2}$ region where $m_{\tilde{\chi}_1^0} \lesssim 10$, the relic density is mainly suppressed through $\chi$-annihilation to fermions through sfermion exchange (low $m_0$), and to $W, Z$ pairs (high $m_0$). This region is excluded primarily by Higgs mass requirements. Along the $\tilde{\chi}_1^0$-LSP boundary the relic density is reduced by $\chi - \tilde{\tau}$-coannihilation, since the coannihilation cross-section is significantly enhanced due to mass degeneracy between the lightest stau and the lightest neutralino. The middle region in the mass plane, the well known Higgs funnel [51], corresponds to high $\tan\beta$ models with $m_{\tilde{\chi}_1^0} \sim 1/2 m_{H^0, A^0}$, giving an increase in $\chi - \chi$-annihilation through

heavy neutral higgs bosons, ($H^0$,$A^0$). The preference for $A_0 \sim 0$ arises mainly from the fact that large parts of the low $m_{1/2}$ regions exhibit charged LSP or tachyonic particles for large values of $|A_0|$, as off-diagonal terms in the third generation sfermion mass matrices grow with $|A_0|$. The preference for high $\tan\beta$ is in part due to the additional relic density suppression through the Higgs channel $\chi$-annihilation.

- **Rare Decays $\mathbf{\Gamma(B_s \to \mu\mu) < 4.5 \cdot 10^{-9}, \Gamma(b \to s + \gamma) \in [3,4] \cdot 10^{-4}}$**
  Of the constraints on decays, $B_s \to \mu\mu$ poses the most stringent one, as the SUSY contribution grows like $\tan\beta^6$. This branching fraction tends to get too large at low values of $m0$ and $m_{1/2}$. The size of the excluded area in the mass plane increases with increasing $tan\beta$ and decreasing $|A_0|$. $\Gamma(b \to s+\gamma)$ is generally too low compared to the central experimental value of $3.55 \cdot 10^{-4}$ and excludes large parts of the low $m_{1/2} \lesssim 500$ range, stretching as far as $m_0 \sim 2000$ for high values of $\tan\beta$ and low $|A_0|$. Too high $\Gamma(B_s \to \mu\mu)$ and too low $\Gamma(b \to s+\gamma)$, together with the requirement of non-tachyonic sparticles constrains the lowest allowed values of $m_{1/2}$.




Theoretical constraints | Th.C+$m_{h0}$ | Th.C+$\Omega h^2$ | Th.C+$\Gamma_{B_s \to \mu\mu, b \to s\gamma}$

**Figure 8**: 2D-histograms in $m_0, m_{1/2}$ and $A_0, \tan\beta$-planes showing the effects of different constraints. The constraints used corresponds to requirements chosen for the initial sample given in table 9

The properties of selected high likelihood models are presented in the results section, 3.1. The relatively wide range of values for SUSY masses and values of $\tan\beta$ for the selected models lead to a wide range of values of phenomenological properties such as average $\not{E}_T$,the average missing energy per SUSY event, $p_T(\tau_1), p_T(\text{jet}_1)$, the average $p_T$ of the leading $\tau$ and the leading jet, and $n_\tau, n_{\text{jet}}$, the average number of $\tau$'s/jets per SUSY event, see 4. In order to construct reference models that cover these different phenomenological properties the sample was clustered according to the phenomenological observables listed above. In order to avoid bias from the scale of the different variables, each variable $x$ is first transformed as $x' = (x_i - \bar{x})/\sigma_x$ so that the mean $\bar{x}' = 0$ and variance $\sigma^2_{x'} = 1$. Because

the non-Gaussian nature of clusters the *g-means* algorithm often fail and split too often. To remedy this the constraining parameters mentioned in section 2.4 are used. By setting an approximate maximum number of possible clusters $n_{\max}$, one gets $\min_P = \lceil N_{\mathrm{OK}}/n_{\max} \rceil$ for the minimum number of points in a cluster and $\min_s = \lceil \log_2 n_{\max} \rceil$ for the maximal splitting depth. The maximal number of iterations per split attempt was set to $\max_i = 10$. Here the number is chosen to be well above the final number of clusters but low enough, for this case $n_{\max} = 40$ was found to be appropriate. The minimal cluster distance parameter was set to $\min_d = 1.3$. The optimization was run $n_{\mathrm{avg}} = 5$ times and an average of 8.8 clusters were found. Thus, one of the results with nine clusters was picked at random. The properties of models at the centroids of these clusters, which can be seen as reference models for search optimization, are presented in the results section 3.1.

## References


[1] H. P. Nilles, *Supersymmetry, Supergravity and Particle Physics*, *Phys.Rept.* **110** (1984) 1–162.

[2] H. E. Haber and G. L. Kane, *The Search for Supersymmetry: Probing Physics Beyond the Standard Model*, *Phys.Rept.* **117** (1985) 75–263.

[3] S. P. Martin, *A Supersymmetry primer*, hep-ph/9709356.

[4] **Particle Data Group** Collaboration, K. Nakamura et al., *Review of particle physics*, *J.Phys.G* **G37** (2010) 075021.

[5] J. R. Ellis, T. Falk, G. Ganis, K. A. Olive, and M. Srednicki, *The CMSSM parameter space at large tan beta*, *Phys.Lett.* **B510** (2001) 236–246, [hep-ph/0102098].

[6] A. H. Chamseddine, R. Arnowitt, and P. Nath, *Locally supersymmetric grand unification*, *Phys. Rev. Lett.* **49** (Oct, 1982) 970–974.

[7] M. Drees and M. M. Nojiri, *The Neutralino relic density in minimal $N = 1$ supergravity*, *Phys.Rev.* **D47** (1993) 376–408, [hep-ph/9207234].

[8] J. R. Ellis, T. Falk, K. A. Olive, and M. Schmitt, *Constraints on neutralino dark matter from LEP-2 and cosmology*, *Phys.Lett.* **B413** (1997) 355–364, [hep-ph/9705444].

[9] A. Lipniacka, *Can SUSY be found at the Tevatron run II?*, hep-ph/0112280.

[10] H. Baer, A. Mustafayev, S. Profumo, A. Belyaev, and X. Tata, *Direct, indirect and collider detection of neutralino dark matter in SUSY models with non-universal Higgs masses*, *JHEP* **0507** (2005) 065, [hep-ph/0504001].

[11] H. Baer, A. Mustafayev, S. Profumo, A. Belyaev, and X. Tata, *Neutralino cold dark matter in a one parameter extension of the minimal supergravity model*, *Phys.Rev.* **D71** (2005) 095008, [hep-ph/0412059].

[12] M. Stoye, *Susy results from cms*, Tech. Rep. CMS-CR-2012-118. CERN-CMS-CR-2012-118, CERN, Geneva, May, 2012.

[13] R. Bruneliere, *Search for supersymmetry at atlas*, Tech. Rep. ATL-PHYS-PROC-2012-073, CERN, Geneva, May, 2012.

[14] O. Buchmueller, R. Cavanaugh, M. Citron, A. De Roeck, M. Dolan, et al., *The CMSSM and NUHM1 in Light of 7 TeV LHC, $B_s$ to mu+mu- and XENON100 Data*, arXiv:1207.7315.

[15] A. Fowlie, M. Kazana, K. Kowalska, S. Munir, L. Roszkowski, et al., *The CMSSM Favoring New Territories: The Impact of New LHC Limits and a 125 GeV Higgs*, arXiv:1206.0264.

[16] S. Ovyn, X. Rouby, and V. Lemaitre, *DELPHES, a framework for fast simulation of a generic collider experiment*, arXiv:0903.2225.

[17] J. Conway, "PGS: Pretty Good Simulator." http://physics.ucdavis.edu/~conway/research/software/pgs/pgs4-general.htm.

[18] G. Hamerly and C. Elkan, *Learning the k in k-means*, in *In Neural Information Processing Systems*, p. 2003, MIT Press, 2003.

[19] O. Buchmueller et al., "The MasterCode Project." http://mastercode.web.cern.ch/mastercode/index.php.

[20] **ATLAS Collaboration** Collaboration, G. Aad et al., *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, *Phys.Lett.* **B716** (2012) 1–29, [arXiv:1207.7214].

[21] A. Gelman, S. Brooks, G. Jones, and X. Meng, *Handbook of Markov Chain Monte Carlo*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press, 2010.

[22] B. Allanach, C. Balazs, G. Belanger, M. Bernhardt, F. Boudjema, et al., *SUSY Les Houches Accord 2*, *Comput.Phys.Commun.* **180** (2009) 8–25, [arXiv:0801.0045].

[23] F. E. Paige, S. D. Protopopescu, H. Baer, and X. Tata, *ISAJET 7.69: A Monte Carlo event generator for p p, anti-p p, and e+ e- reactions*, hep-ph/0312045.

[24] S. Heinemeyer, W. Hollik, and G. Weiglein, *FeynHiggs: A Program for the calculation of the masses of the neutral CP even Higgs bosons in the MSSM*, *Comput.Phys.Commun.* **124** (2000) 76–89, [hep-ph/9812320].

[25] P. Bechtle, O. Brein, S. Heinemeyer, G. Weiglein, and K. E. Williams, *HiggsBounds: Confronting Arbitrary Higgs Sectors with Exclusion Bounds from LEP and the Tevatron*, *Comput.Phys.Commun.* **181** (2010) 138–167, [arXiv:0811.4169].

[26] P. Gondolo et al., *DarkSUSY: Computing supersymmetric dark matter properties numerically*, *JCAP* **0407** (2004) 008, [astro-ph/0406204].

[27] K. Hagiwara et al., *Review of Particle Physics*, *Physical Review D* **66** (2002) 010001+.

[28] **DELPHI Collaboration** Collaboration, J. Abdallah et al., *Searches for supersymmetric particles in e+ e- collisions up to 208-GeV and interpretation of the results within the MSSM*, *Eur.Phys.J.* **C31** (2003) 421–479, [hep-ex/0311019].

[29] T. Sjostrand, S. Mrenna, and P. Z. Skands, *A Brief Introduction to PYTHIA 8.1*, *Comput.Phys.Commun.* **178** (2008) 852–867, [arXiv:0710.3820].

[30] **LHCb Collaboration** Collaboration, R. Aaij et al., *Search for the rare decays Bs to mu+ mu- and B0 to mu+ mu-*, *Phys.Lett.* **B708** (2012) 55–67, [arXiv:1112.1600].

[31] **WMAP** Collaboration, E. Komatsu et al., *Seven-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation*, *Astrophysbuch. J. Suppl.* **192** (2011) 18, [arXiv:1001.4538].

[32] **Heavy Flavor Averaging Group** Collaboration, D. Asner et al., *Averages of b-hadron, c-hadron, and $\tau$-lepton Properties*, arXiv:1010.1589.

[33] P. Gondolo et al., "DarkSUSY darksusy-5.0.5." Manual and short description of routines, June, 2009.

[34] B. Allanach and C. Lester, *Multi-dimensional mSUGRA likelihood maps*, *Phys.Rev.* **D73** (2006) 015013, [hep-ph/0507283].

[35] **ALEPH Collaboration, DELPHI Collaboration, L3 Collaboration, OPAL Collaborations, LEP Working Group for Higgs Boson Searches** Collaboration, S. Schael et al., *Search for neutral MSSM Higgs bosons at LEP*, *Eur.Phys.J.* **C47** (2006) 547–587, [hep-ex/0602042].

[36] I. Niessen, *Supersymmetric Phenomenology in the mSUGRA Parameter Space*, arXiv:0809.1748. Master Thesis (Advisors: Wim Beenakker, Nicolo de Groot).

[37] W. K. Hastings, *Monte carlo sampling methods using markov chains and their applications*, *Biometrika* **57** (1970), no. 1 97–109.

[38] R. V. Craiu and A. F. D. Narzo, *A mixture-based approach to regional adaptation for mcmc*, tech. rep., 2009.

[39] Y. Guan, R. Fleißner, P. Joyce, and S. M. Krone, *Markov chain monte carlo in small worlds*, *Statistics and Computing* **16** (June, 2006) 193–202.

[40] M. West, *Approximating posterior distributions by mixtures*, *Journal of the Royal Statistical Society. Series B (Methodological)* **55** (1993), no. 2 409–422.

[41] R. Gelman, , G. O. Roberts, A. Gelman, and W. R. Gilks, *Weak convergence and optimal scaling of random walk metropolis algorithms*, 1994.

[42] J. MacQueen, "Some methods for classification and analysis of multivariate observations.." Proc. 5th Berkeley Symp. Math. Stat. Probab., Univ. Calif. 1965/66, 1, 281-297 (1967)., 1967.

[43] J. L. Bentley, *Multidimensional binary search trees used for associative searching*, *Commun. ACM* **18** (Sept., 1975) 509–517.

[44] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. Piatko, R. Silverman, and A. Y. Wu, *An efficient k-means clustering algorithm: Analysis and implementation*, 2000.

[45] D. Arthur and S. Vassilvitskii, *k-means++: The advantages of careful seeding*, Technical Report 2006-13, Stanford InfoLab, June, 2006.

[46] M. A. Stephens, *Edf statistics for goodness of fit and some comparisons*, *Journal of the American Statistical Association* **69** (1974), no. 347 730–737.

[47] S. Profumo, *The Quest for Supersymmetry: Early LHC Results versus Direct and Indirect Neutralino Dark Matter Searches*, *Phys.Rev.* **D84** (2011) 015008, [arXiv:1105.5162].

[48] J. R. Ellis, T. Falk, K. A. Olive, and M. Srednicki, *Calculations of neutralino-stau coannihilation channels and the cosmologically relevant region of MSSM parameter space*, *Astropart.Phys.* **13** (2000) 181–213, [hep-ph/9905481].

[49] H. Baer, E.-K. Park, and X. Tata, *Collider, direct and indirect detection of supersymmetric dark matter*, *New J. Phys.* **11** (2009) 105024, [arXiv:0903.0555].

[50] H. Baer, T. Krupovnickas, A. Mustafayev, E.-K. Park, S. Profumo, et al., *Exploring the BWCA (bino-wino co-annihilation) scenario for neutralino dark matter*, *JHEP* **0512** (2005) 011, [hep-ph/0511034].

[51] L. Roszkowski, R. Ruiz de Austri, and T. Nihei, *New cosmological and experimental constraints on the CMSSM*, *JHEP* **0108** (2001) 024, [hep-ph/0106334].